Chapter 10

Statistical Graphs

Learning Objectives

After completing this chapter, you will be able to

- know the existence of different types of data including discrete data and continuous data.
- understand the criteria and ways of organizing data.
- construct and interpret histograms and scatter diagrams.
- identify the difference between histograms and bar charts.
- compare the presentations of a given set of data by using various graphs.
- choose suitable graphs to present a given set of data.
- construct histograms and scatter diagrams using computer software.

The table shows the neck and waist sizes of 20 people.

Neck size (cm)	Waist size (cm)	Neck size (cm)	Waist size (cm)
43.2	86.5	35.2	68
40.6	81	38.2	76
33.0	63.5	41.6	80.5
39.4	83	36	72.5
30.6	61	30	60.5
31.8	63.5	42	85.5
35.6	76	37.4	73
39.4	77	33.6	68.5
36.8	76	35	73
34.4	70	38.8	74

Describe the relation between the neck and waist sizes.

It is given that the neck size of Kenny is 45 cm. What is his approximate waist size? Why?

Preview

[Basic knowledge required for this chapter.]

Basic Knowledge

1. Statistics is a branch of mathematics studying how data is collected, organized, presented and analyzed.

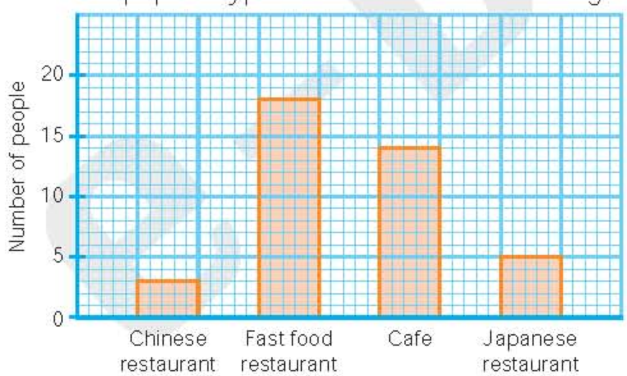
2. Frequency distribution table

Age	Tally	Frequency
12	~~////~~ ///	8
13	~~////~~ ~~////~~ ~~////~~ /	16
14	~~////~~ ~~////~~	10
15	////	4

3. Statistical graphs

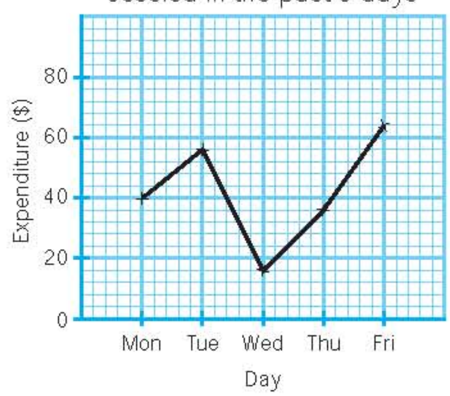
(a) Bar chart

Most popular types of restaurants with teenagers



(b) Broken-line graph

Daily expenditure of Jessica in the past 5 days



(c) Pie chart

Average daily usage of plastic bags among 60 students

(d) Stem-and-leaf diagram

Prices of 20 magazines at a news-stand

Stem ($10)	Leaf ($1)
0	8 8
1	0 0 0 2 2 5 5 8
2	0 0 0 5 5
3	6 8
4	0 0 5

10.1 Types of Data

In Chapter 6 of S1A, we have learned the basic steps for doing statistical work and using simple methods to collect, organize and present data. Now, we are going to discuss different types of data.

Class Activity 10.1

Aim: To learn different types of data

According to the properties of the following data, put a '✓' in the appropriate box.

Show Ans.

	Exact value	Approximate value	Obtained by counting	Obtained by measurement
Ⓐ  Number of pens				
Ⓑ  Height of a student				
Ⓒ  Time spent on dinner				
Ⓓ  Weight of a school bag				
Ⓔ  Number of family members				
Ⓕ  Amount of money in a wallet				

Now I see ...

Data obtained by counting are exact values. Data obtained by measurement are approximate values.

From Class Activity 10.1, we find that the data are of two different natures.

For Ⓐ and Ⓔ, number of pens and number of family members can be 1, 2, 3, ... etc. For Ⓕ, amount of money in a wallet can be \$32, \$45, \$48.7, \$53.4, ... etc. The value of each datum is a specific value. This type of data is known as **discrete data**.

For Ⓑ, the height of a student can be 168 cm or 170.21 cm or even 174.214 69 ... cm by measurement. The same goes for Ⓒ and Ⓓ. The value of each datum can be any value within a range. This type of data is known as **continuous data**.

In general, data obtained by measurement such as length, weight and time are continuous data.

Extension 10.1

Solution

Determine the type of each of the following sets of data collected, and put a '✓' in the appropriate box.

	Discrete data	Continuous data
(a) Number of days with very hot weather warnings each year	☐	☐
(b) Number of heartbeats per minute	☐	☐
(c) Weights of students	☐	☐
(d) Number of visitors in Ocean Park per day	☐	☐
(e) Travel time by a bus between two bus stops	☐	☐
(f) Daily rainfall of June	☐	☐

10.2 Grouping of Data

If the volume of data collected is large and widespread, it is better to group them into classes for further analysis.

discrete data 離散數據　　continuous data 連續數據

A Frequency distribution

In order to improve the quality of services, New City Bus Company carried out a survey about the bus services in Central. Figure 10.1 shows the waiting time of 100 passengers (correct to the nearest minute).

20	13	20	17	23	15	12	26	20	19
25	11	6	25	16	22	31	8	20	10
21	18	28	21	30	6	18	29	16	14
33	8	13	19	15	12	13	16	22	13
15	19	15	15	20	25	20	26	19	18
19	20	19	14	12	6	17	7	21	25
6	35	12	22	26	31	18	24	30	15
20	27	14	17	18	8	14	9	7	22
12	32	25	7	23	20	23	19	14	25
22	15	28	19	12	25	16	21	32	19

Figure 10.1

The data collected are large in volume and distributed extensively. To analyze their distribution, we can group them into classes as follows:

1. Read the data carefully and find the largest and smallest values.
2. Determine the number of classes and the range of each class to cover all data.
3. Use tallies to record the number of data in each class.

A suitable number of classes can help us show the trend of the distribution of data.

According to Figure 10.1, the smallest value is 6 minutes and the largest value is 35 minutes. If the range of each class is 5 minutes, a frequency distribution table can be constructed as follows:

Waiting time (range of each class = 5 min)

Waiting time (min)	Tally	Frequency
6 - 10	卌 卌 //	12
11 - 15	卌 卌 卌 卌 ///	23
16 - 20	卌 卌 卌 卌 卌 卌	30
21 - 25	卌 卌 卌 卌	20
26 - 30	卌 ////	9
31 - 35	卌 /	6

Table 10.1

From the table above, we find that most of the passengers have to wait between 11 minutes to 25 minutes.

B Concepts about classes

There are 6 classes in Table 10.1 in the last section. Each class can be expressed as a **class interval**.

e.g. The second class interval in the table is '11 min - 15 min'.

We will study some concepts about classes by taking the class '11 min - 15 min' as an example.

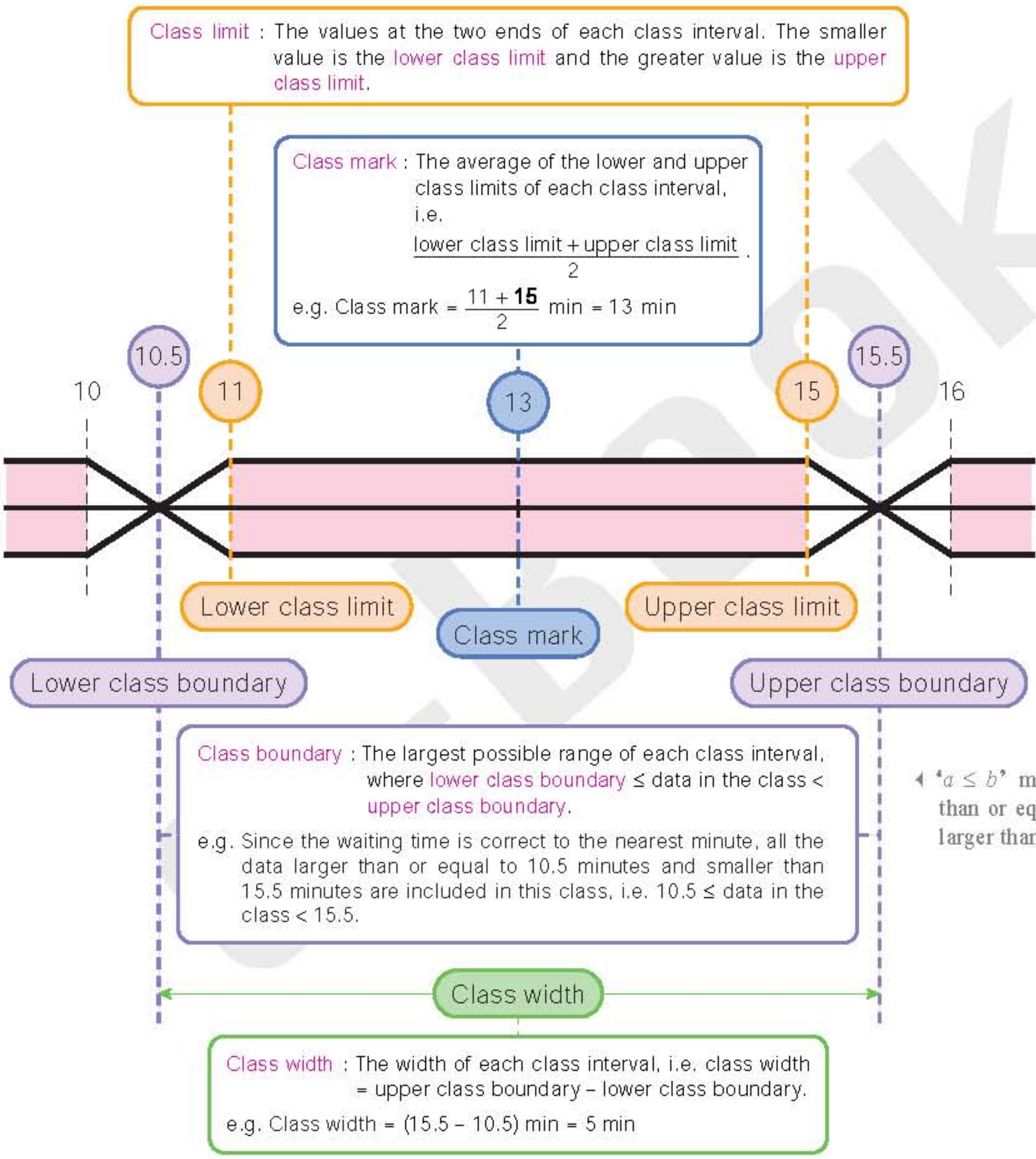


Figure 10.2

'$a \le b$' means 'a is smaller than or equal to b', i.e. 'b is larger than or equal to a'.

Note: **The upper class boundary of a class interval is also the lower class boundary of the next class interval.**

class interval 組區間
upper class limit 上組限
lower class boundary 下組界
class limit 組限
class mark 組中點
upper class boundary 上組界
lower class limit 下組限
class boundary 組界
class width 組距

Example 10.1 Interpreting the frequency distribution table

Additional Example 10.1

The frequency distribution table shows the ages of the students in a painting class.

(a) What is the frequency of the second class interval?
(b) Find the upper and lower class limits of the third class interval.
(c) Find the class mark of the fourth class interval.
(d) Find the upper and lower class boundaries of the sixth class interval.
(e) Which class interval has the most students? What percentage of students are in it?

Age	Frequency
20 - 29	1
30 - 39	3
40 - 49	9
50 - 59	12
60 - 69	10
70 - 79	5
Total	40

Solution

(a) The frequency of the second class interval is $\underline{\underline{3}}$.

(b) For the third class interval, the upper class limit is $\underline{\underline{49}}$ and the lower class limit is $\underline{\underline{40}}$.

(c) For the fourth class interval, the class mark $= \dfrac{50+59}{2}$
$= \underline{\underline{54.5}}$

(d) For the sixth class interval, the upper class boundary is $\underline{\underline{79.5}}$ and the lower class boundary is $\underline{\underline{69.5}}$.

(e) Most students are in the class interval 50 - 59.

$\because$ The number of students in this class interval = 12

$\therefore$ Required percentage $= \dfrac{12}{40} \times 100\%$
$= \underline{\underline{30\%}}$

Classwork 10.1

Solution

(a) The following table shows the weights of 50 people. Complete the table.

Weight (kg)	Lower class boundary (kg)	Upper class boundary (kg)	Class mark (kg)	Frequency
45 - 49	44.5	49.5		2
50 - 54				9
55 - 59				16
60 - 64				14
65 - 69				6
70 - 74				3

(b) What is the class width?
(c) Which class interval has the fewest people in it?

Skills Upgrading Corner 10.1

Solution

The following are the average lap times (correct to the nearest second) of the racing cars in a race.

234	221	242	254	243	238	242	239
227	229	263	236	234	235	240	239
241	256	250	229	230	239	264	248
252	247	265	246	232	234	257	228

(a) According to the data above, complete the following table.

Lap time (s)	Class boundaries (s)	Class mark (s)	Tally	Frequency
220 - 229				
			Total	

(b) Which class interval has the highest frequency?

(c) Which class interval does each of the following lap times fall into?

(i) 234.5 s **(ii)** 259.5 s

(d) What percentage of the racing cars take less than 239.5 s to complete a lap?

Exercise 10A

Level 1

1. The following table shows the daily wages of workers in a factory.

Daily wages ($)	250 - 299	300 - 349	350 - 399	400 - 449	450 - 499	500 - 549
Frequency	5	8	14	17	20	6

(a) Which class interval has the most workers?

(b) How many workers are there in the factory?

(c) For the fourth class interval, find

(i) the upper class limit.

(ii) the lower class limit.

(iii) the class mark.

2. The following table shows the weights of 50 bags of vegetables.

Weight (kg)	0.2 - 0.5	0.6 - 0.9	1.0 - 1.3	1.4 - 1.7	1.8 - 2.1	2.2 - 2.5
Frequency	4	7	14	17	5	3

(a) Find the upper and lower class boundaries of the fifth class interval.
(b) What is the class width?
(c) Find the number of bags of vegetables with weights less than 1.35 kg.

3. The following table shows the heights of a group of girls.

Height (cm)	140 - 144	145 - 149	150 - 154	155 - 159	160 - 164
Frequency	9	14	46	30	11

(a) Which class interval does the class mark 157 cm fall into?
(b) Which class interval does each of the following heights fall into?
   (i) 146 cm
   (ii) 154.5 cm
(c) Find the number of girls with heights of 149.5 cm or more.

4. The following table shows the weights of 50 MP3 players in Melody Audio Video Shop.

Weight (g)	30 - 49	50 - 69	70 - 89	90 - 109	110 - 129	130 - 149	150 - 169
Frequency	2	5	9	12	18	3	1

(a) What is the class width?
(b) Which class interval does the upper class boundary 109.5 g belong to?
(c) Find the percentage of MP3 players which weigh less than 89.5 g.

Level 2

5. **(a)** The following table shows the average time spent by 40 students in watching television. Complete the table.

Time (min)	Class boundaries (min)	Class mark (min)	Frequency
1 - 30	0.5 - 30.5		3
31 - 60			6
61 - 90			12
91 - 120			
121 - 150			6
151 - 180			7

(b) Which class interval has the fewest students?

6. The following are the waiting time of 45 customers before getting their seats in ABC Sushi Restaurant, correct to the nearest minute.

76	68	46	72	59	55	53	58	38
52	55	81	54	45	34	80	43	51
40	61	59	45	58	61	73	62	49
55	53	66	56	47	48	42	51	56
29	48	39	64	67	74	65	65	42

(a) According to the above data, complete the following table.

Waiting time (min)	Class boundaries (min)	Class mark (min)	Tally	Frequency
26 - 35				
76 - 85				

(b) What is the class width?

(c) Find the number of customers who have to wait for less than 65.5 min.

7. The following are the weights of 30 boxes of nuts, correct to the nearest 0.1 g.

150.3	149.8	150.2	148.2	151.4	151.2	150.1	148.9	148.7	150.8
149.6	149.5	148.6	150.1	150.2	150.3	151.2	150.2	149.9	150.0
148.0	148.3	150.4	150.1	148.7	149.0	150.6	151.1	149.4	149.9

(a) Among the class widths of 0.2 g, 0.5 g and 1 g, which is the most suitable class width for the above data?

(b) According to the class width chosen in **(a)**, construct a frequency distribution table and list the class boundaries and class mark of each class.

8. The following stem-and-leaf diagram shows the Chinese composition scores of a group of students.

Chinese composition scores of a group of students

Stem (tens)	Leaf (units)
3	3 7
4	0 0 1 2 2 2 3 5 5 6 7 8 8 9
5	1 1 2 3 3 3 4 4 5 5 5 6 6 6 7 8 8 9 9
6	0 1 2 4 5 5 6 7 8
7	2 3 4 5 5 7

(a) Using 30 - 34 as the first class interval, construct a frequency distribution table and list the class boundaries and class mark of each class.

(b) How many students are there in the group?

(c) Which class interval has the most students? How many students are there?

(d) Find the percentage of students with scores of 59.5 or more.

10.3 Histograms

A **histogram** is a common statistical graph. Data are represented by rectangular bars with no gaps in between. The height of each rectangular bar represents the frequency of each class while the width represents the class width.

e.g. According to the data in Table 10.1 (page 10.5), we can obtain Table 10.2 and the histogram in Figure 10.3.

Waiting time (min)	Class boundaries (min)	Class mark (min)	Frequency
6 - 10	5.5 - 10.5	8	12
11 - 15	10.5 - 15.5	13	23
16 - 20	15.5 - 20.5	18	30
21 - 25	20.5 - 25.5	23	20
26 - 30	25.5 - 30.5	28	9
31 - 35	30.5 - 35.5	33	6

Table 10.2

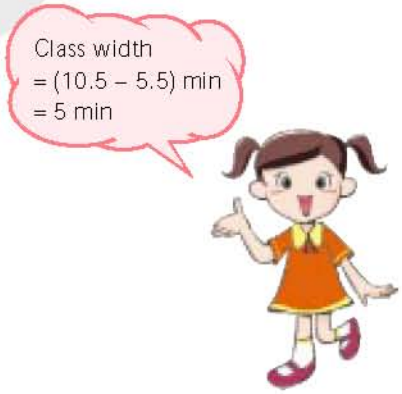


Figure 10.3

histogram 組織圖

I.T. Tool

[Refer to page 10.36]

The frequency distribution can be read from a histogram. Thus a histogram can help us understand the distribution of the data more easily.

When drawing a histogram, the following should be noted.

1. A title should be given.
2. Label 'Frequency' on the vertical axis.
3. Each bar is perpendicular to the horizontal axis, and there should be no gaps between the bars.
4. Only the class mark is used to represent each class.

A bar chart is very similar to a histogram. However, they have the following differences.

1. There are gaps between bars in a bar chart but not in a histogram.
2. Some data can be represented by bar charts but not by histograms. For example, the population of each area in a country or the number of customers for different products etc. In general, continuous data can be represented by histograms while discrete data can be represented by bar charts.

Extension 10.2

Solution

Which of the following data can be represented by a histogram?

(a) The sales record of TV sets in the past 10 years
(b) The number of people visiting Space Museum every month
(c) The age distribution of local citizens
(d) The heights of students in a school
(e) The time taken by a class of students to complete a lap of the playground

Example 10.2 Constructing a histogram

Additional Example 10.2

The table shows the high jump records of S2A students.

Height (cm)	Frequency
120 - 124	2
125 - 129	4
130 - 134	9
135 - 139	10
140 - 144	8
145 - 149	4
150 - 154	3

(a) Construct a frequency distribution table with class boundaries and class marks.

(b) Based on the frequency distribution table in **(a)**, construct a histogram.

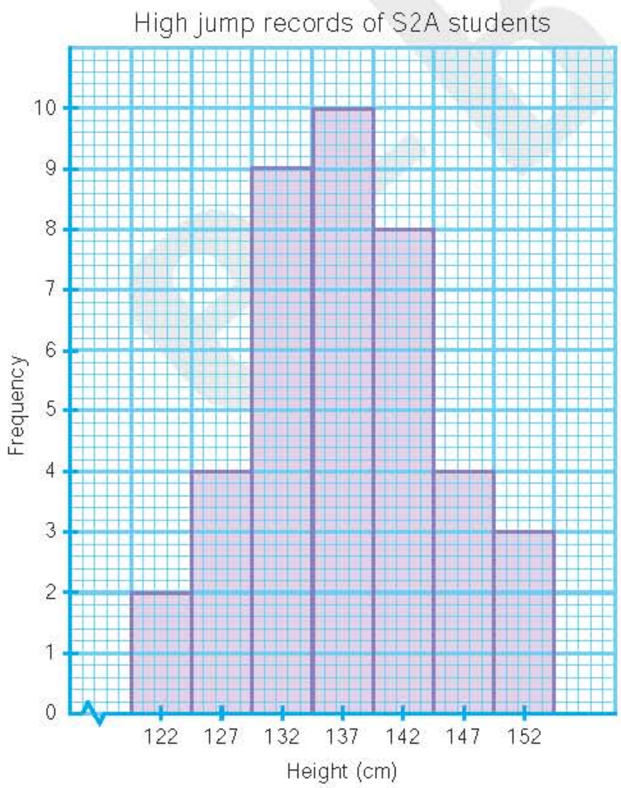
(c) Any students with records reaching 144.5 cm can join the open competition. How many students have met the requirement?

Solution

(a)

Height (cm)	Class boundaries (cm)	Class mark (cm)	Frequency
120 - 124	119.5 - 124.5	122	2
125 - 129	124.5 - 129.5	127	4
130 - 134	129.5 - 134.5	132	9
135 - 139	134.5 - 139.5	137	10
140 - 144	139.5 - 144.5	142	8
145 - 149	144.5 - 149.5	147	4
150 - 154	149.5 - 154.5	152	3

(b) [Analysis: Class width = (124.5 − 119.5) cm = 5 cm
$\therefore$ Both the differences between class marks and the widths of the bars are 5 (or multiples of 5) units.]



(c) [Analysis: Since 144.5 cm is the upper class boundary of the fifth class interval, the frequencies of the 6th and the 7th class intervals should be considered.]

$3 + 4 = \underline{\underline{7}}$ students have met the requirement.

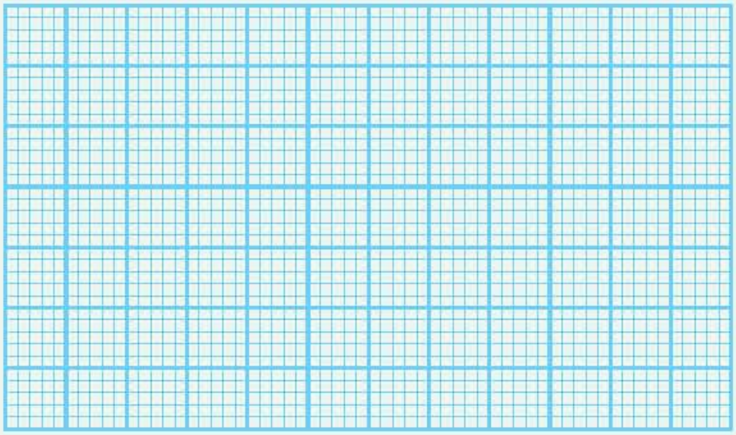
Classwork 10.2

Solution

(a) The following table shows the heights of 50 plants. Complete the table.

Height (cm)	Class boundaries (cm)	Class mark (cm)	Frequency
10 - 14			6
15 - 19			10
20 - 24			19
25 - 29			13
30 - 34			2

(b) Based on the frequency distribution table in (a), draw a histogram.



(c) Plants with heights of 24.5 cm or more will be transferred to the Wetland Park. How many plants will be transferred?

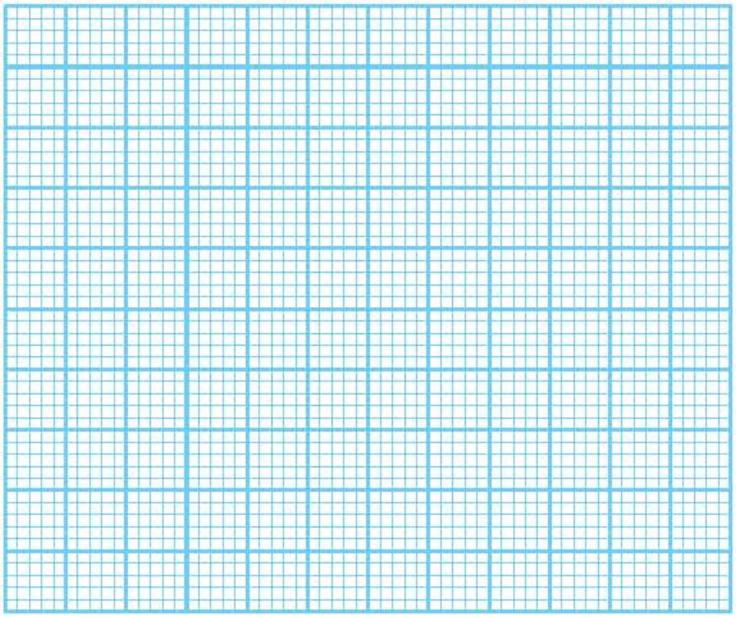
Skills Upgrading Corner 10.2

Solution

(a) The following table shows the ages of members in Active Fitness Centre. Complete the table.

Age	Class boundaries	Class mark	Frequency
15 - 19			12
20 - 24			30
25 - 29			34
30 - 34			18
35 - 39			6
		Total	

(b) Based on the frequency distribution table in **(a)**, draw a histogram.



(c) How many members are between 19.5 and 29.5 years old?

(d) If members below 19.5 years old are called 'youth members', what percentage of the members are 'youth members'?

Exercise 10B

Level 1

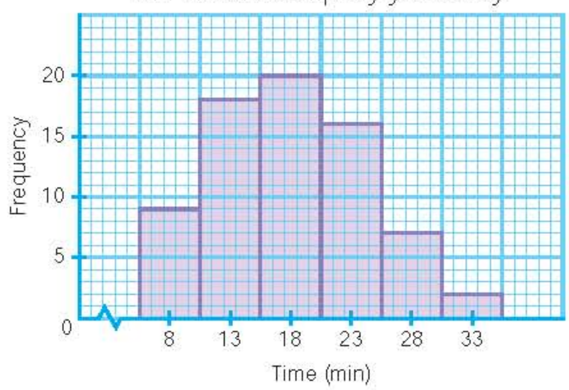
1. The histogram shows the duration of incoming calls of Fine Art Company yesterday. It is known that the first class interval is 6 min - 10 min.

(a) Which class interval has the highest frequency?

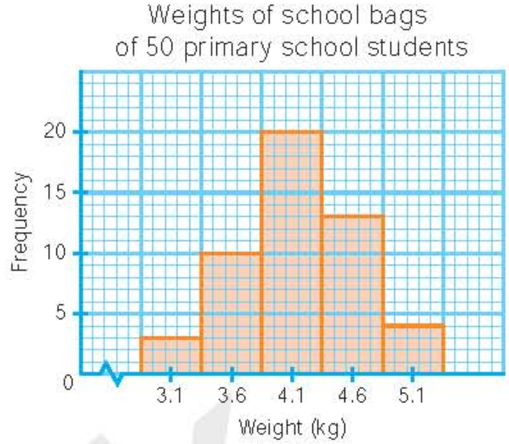
(b) Which class interval has the lowest frequency?

(c) How many incoming calls were received yesterday?

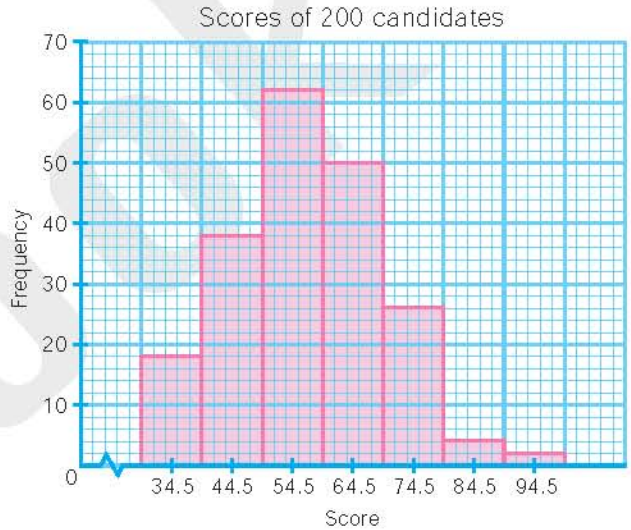
Duration of incoming calls of Fine Art Company yesterday



2. The histogram shows the weights of school bags of 50 primary school students. It is known that the first class interval is 2.9 kg - 3.3 kg.
   (a) How many primary school students have their school bags weighed between 2.85 kg and 3.85 kg?
   (b) If a school bag weighs 3.85 kg or more is regarded as overloaded, how many school bags of these primary school students are overloaded?



3. The histogram shows the scores of 200 candidates in a public examination. It is known that the first class interval is 30 - 39.
   (a) Which class interval has the fewest candidates? How many candidates are there?
   (b) If candidates with scores of 79.5 or above will obtain distinctions, how many candidates will obtain distinctions?
   (c) If candidates with scores below 49.5 will fail, what percentage of the candidates will fail?



Level 2

4. The average time taken for 30 people to have haircut each (in minute) is shown as follows:

51	40	28	33	135	27	35	42	65	40
48	110	70	23	30	90	50	33	25	60
52	86	66	45	125	60	95	100	55	72

   (a) According to the above data, complete the following table.

Time (min)	Class boundaries (min)	Class mark (min)	Tally	Frequency
20 - 39				
40 - 59				
60 - 79				
80 - 99				
100 - 119				
120 - 139				

   (b) Based on the frequency distribution table in **(a)**, draw a histogram.

5. The following table shows the average time 200 employed people spent on exercise each day.

Time (min)	1 - 10	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60	61 - 70
Frequency	12	30	67	37	32	17	5

(a) Construct a frequency distribution table with class boundaries and class marks.

(b) Based on the frequency distribution table in **(a)**, draw a histogram.

(c) According to the suggestion of the Department of Health, an employed person is recommended to exercise for 20.5 min or more every day. What percentage of employed people did not follow the recommendations?

6. The following table shows the scores of 80 people in a psychological test.

Score	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74	75 - 79
Frequency	3	11	9	13	21	6	8	5	4

(a) Construct a histogram to represent the above data.

(b) Any person with the score not less than 64.5 is regarded as a gentle person. How many people are regarded as gentle in this group?

(c) Any person with the score less than 49.5 is regarded as an active person. What percentage of people are active?

(d) Any person with the score between 49.5 and 64.5 is regarded as a thinker. What is the most common type of people within this group?

10.4 Scatter Diagrams

A Introduction to scatter diagrams

The following table shows the scores of 10 S1A students in a Mathematics test and a Science test.

Student	A	B	C	D	E	F	G	H	I	J
Mathematics	84	66	92	44	28	42	76	90	48	34
Science	72	62	82	36	32	48	74	86	46	40

Table 10.3

Can you see any relation between the test scores in Mathematics and Science?

Representing the data above on a rectangular coordinate plane can help in finding out the relations between data. (See Figure 10.4)

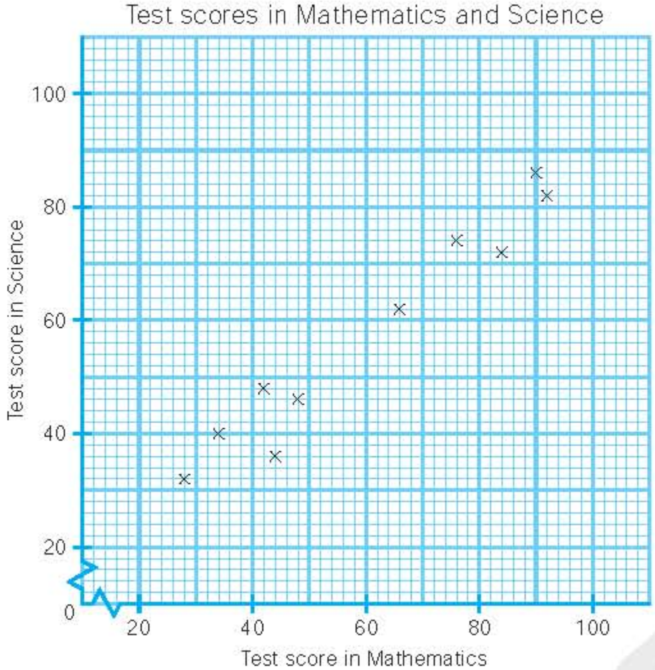


Figure 10.4

[Refer to page 10.38]

In Figure 10.4, each point represents the score of a student, where the x-coordinate and the y-coordinate represent the test scores in Mathematics and Science respectively. This graph is known as a **scatter diagram**, it can show the relation between two sets of data.

e.g. In Figure 10.4, we can see the following relation:

'In this class, the students with low scores in Mathematics also have low scores in Science; conversely, the students with high scores in Mathematics also have high scores in Science.'

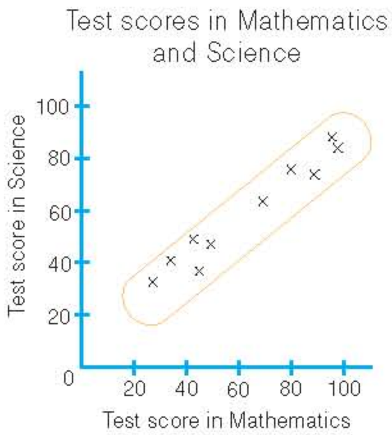
When drawing a scatter diagram, the following should be noted.

1. A title should be given.
2. The x-axis and the y-axis must be properly labelled with items and scales (including units).
3. There should be sufficient number of points (at least 10 points in general) in a scatter diagram.

scatter diagram 散點圖

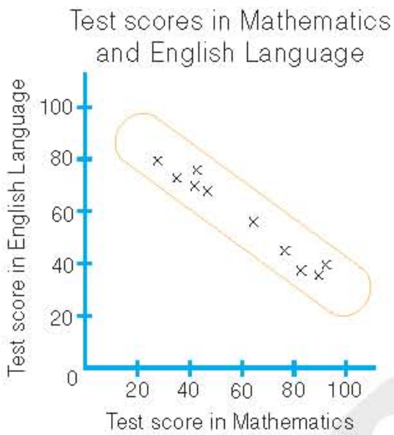
B Relation between data

Figure 10.5(a), (b) and (c) show the scatter diagrams of 'Test scores in Mathematics and Science', 'Test scores in Mathematics and English Language' and 'Test scores in Mathematics and Physical Education' respectively. Can you see any relations between data from these three diagrams?



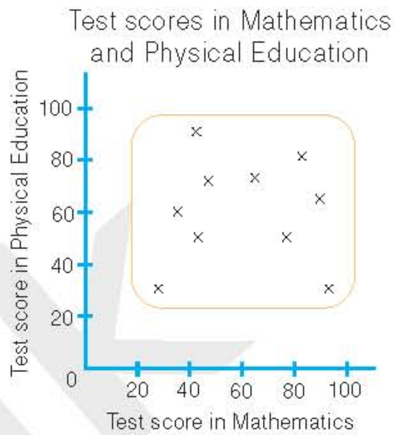
The students with high scores in Mathematics also have high scores in Science. There is a 'positive relation' between these two sets of data.

(a)



The students with high scores in Mathematics tend to have low scores in English Language. There is a 'negative relation' between these two sets of data.

(b)



The scores of students in Mathematics and Physical Education have no clear relation. There is 'no obvious relation' between these two sets of data.

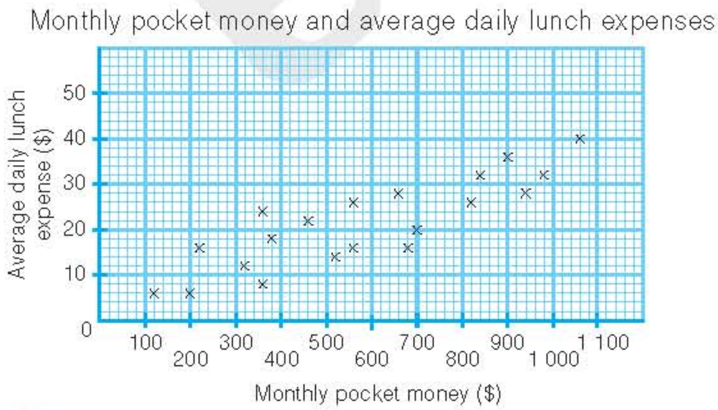
(c)

Figure 10.5

Additional Example 10.3

Example 10.3 Analyzing a scatter diagram with positive relation

Based on the following scatter diagram, describe the relation between the amount of monthly pocket money and the average daily lunch expenses of a group of students.



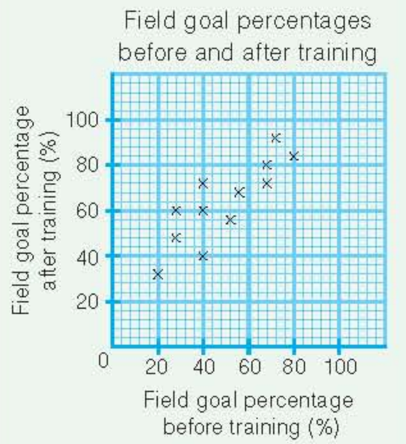
Solution

From the diagram, we can see that the more monthly pocket money the students have, the more their average daily lunch expenses are. There is a positive relation between the amount of their monthly pocket money and average daily lunch expenses.

Classwork 10.3

Solution

The following scatter diagram shows the *field goal percentages* of a group of students before and after training. Describe the relation between these two sets of data.

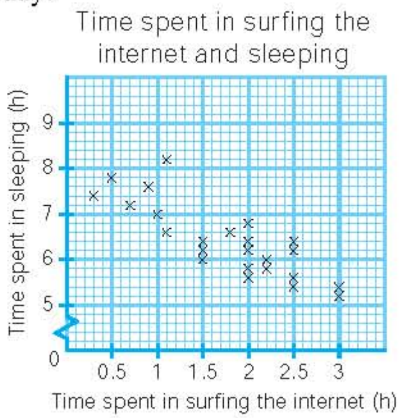


field goal percentage 入球率

Additional Example 10.4

Example 10.4 Analyzing a scatter diagram with negative relation

Based on the following scatter diagram, describe the relation between the time spent by 24 students in surfing the internet and sleeping in a day.



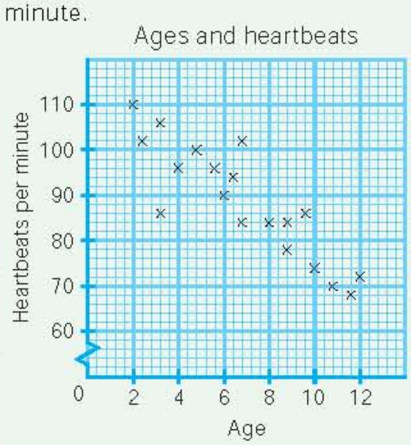
Solution

From the diagram, we can see that for the 24 students, the longer they spend in surfing the internet, the less they sleep. There is a negative relation between their time spent in surfing the internet and sleeping.

Classwork 10.4

Solution

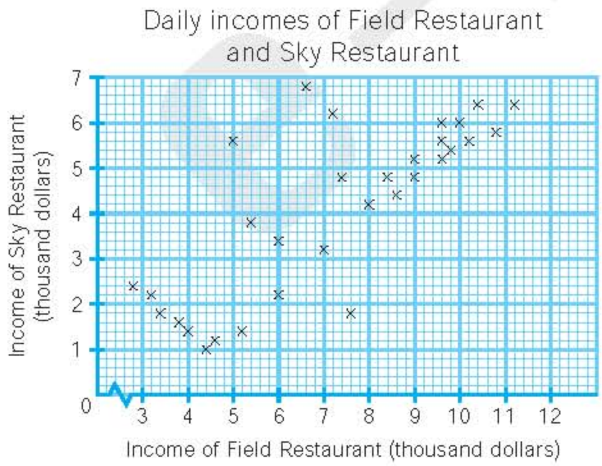
According to the following scatter diagram, describe the relation between the ages of a group of children and their heartbeats per minute.



Skills Upgrading Corner 10.3

Solution

The following scatter diagram shows the daily incomes of two nearby restaurants in March.



(a) What is the relation between the daily incomes of these two restaurants in March?

(b) Overall, which restaurant has higher income?

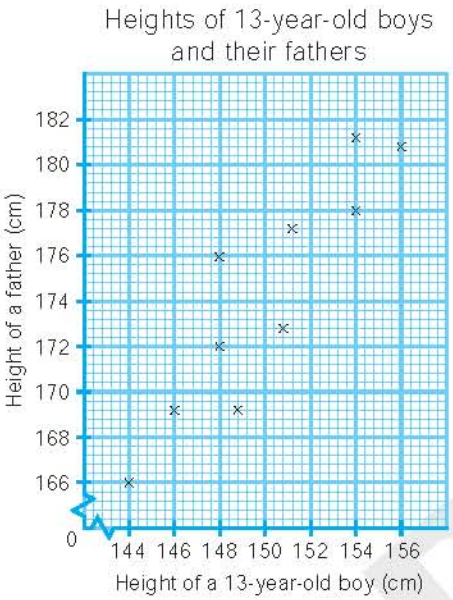
(c) When the daily income of Field Restaurant is above \$8 000, describe the relation between the daily incomes of these two restaurants.

(d) When the daily income of Field Restaurant is below \$5 000, describe the relation between the daily incomes of these two restaurants.

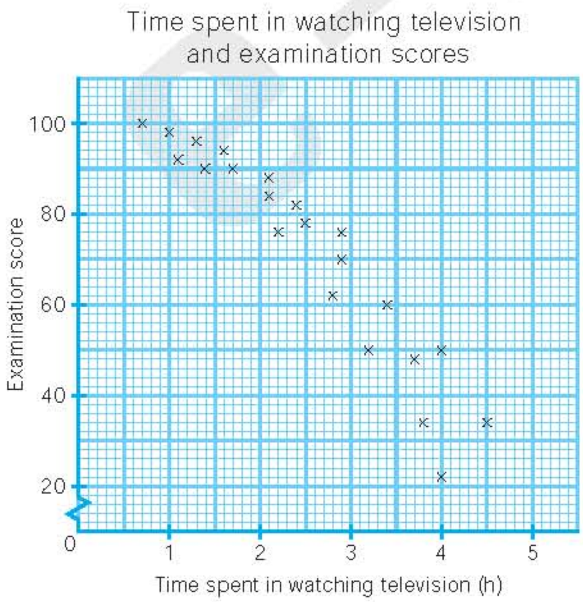
Exercise 10C

Level 1

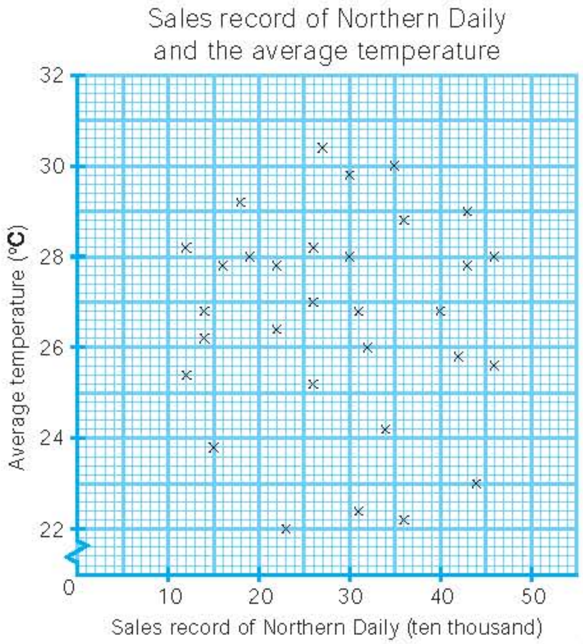
1. According to the scatter diagram, describe the relation between the heights of 13-year-old boys and their fathers.



2. Based on the scatter diagram, describe the relation between the time spent by a group of students in watching television and their examination scores.



3. According to the scatter diagram, describe the relation between the sales record of Northern Daily and the average temperature.



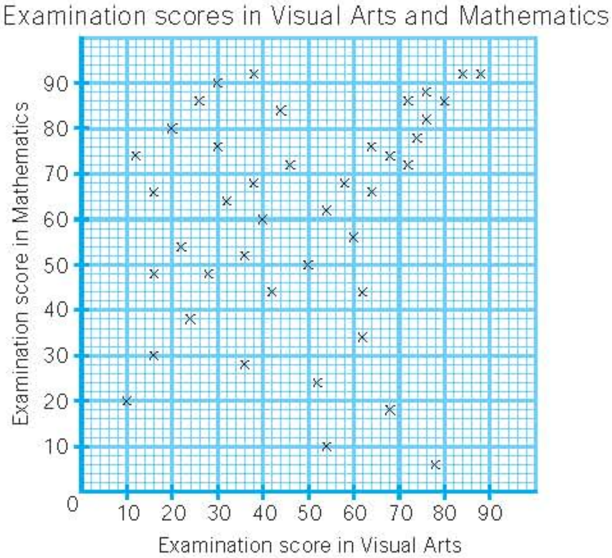
Level 2

4. The scatter diagram shows the monthly salaries of a group of employees and the percentages of their salaries invested.

(a) Did the employees with higher monthly salaries also have higher percentages of salaries invested?

(b) For the employees with monthly salaries higher than $50 000, is it true that the higher their monthly salaries are, the higher the percentages of their salaries are invested?

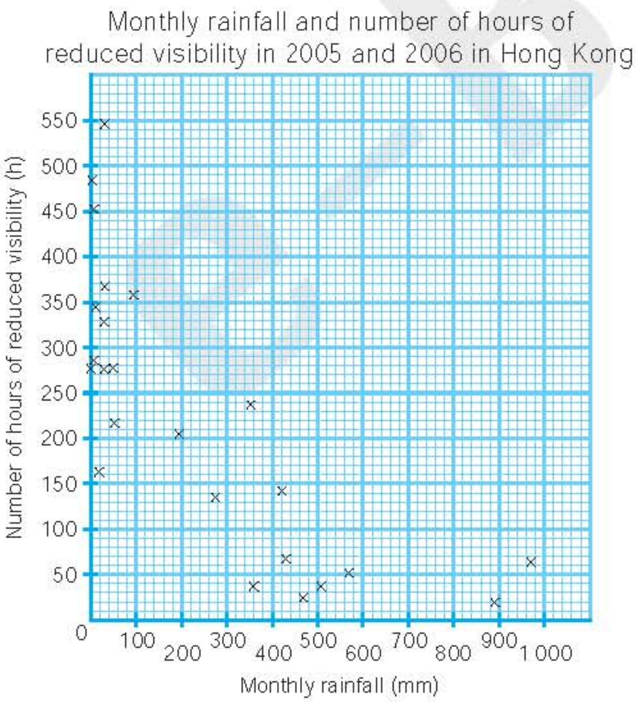
5. The scatter diagram shows the examination scores of S1A students in Visual Arts and Mathematics.



(a) If students with scores of 80 or above will get grade A, how many students will get grade A in each subject?

(b) Among students who got 50 marks or above in both subjects, describe the relation between their scores in Visual Arts and Mathematics.

6. The scatter diagram shows the monthly rainfall and number of hours of reduced visibility in 2005 and 2006 in Hong Kong.



(a) When the monthly rainfall was less than 100 mm, were there long hours of reduced visibility?

(b) For the months with rainfall over 250 mm, describe the relation between the monthly rainfall and the number of hours of reduced visibility.

(c) Determine whether the following statement is reasonable.
'In 2005 and 2006, the visibility was always low in the rainy season.'

7. The following table shows the annual incomes of 20 *strikers* and their goals scored last season.

Striker	A	B	C	D	E	F	G	H	I	J
Annual income (ten thousand dollars)	56	85	64	74	22	17	42	48	16	88
Goals scored	6	25	8	13	5	7	10	13	5	7
Striker	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**
Annual income (ten thousand dollars)	37	41	56	41	35	67	87	36	10	32
Goals scored	2	7	10	1	4	5	11	21	3	3

(a) Draw a scatter diagram to present the data above.

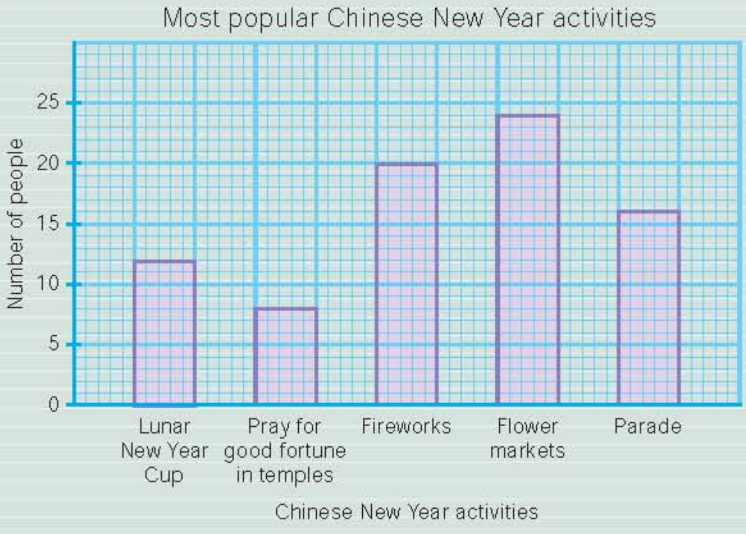
(b) Based on the scatter diagram in **(a)**, describe the relation between the annual incomes of these strikers and their goals scored last season.

10.5 Choosing Appropriate Statistical Graphs

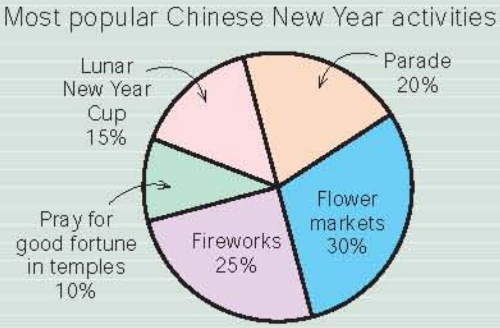
Class Activity 10.2

Aim: To choose an appropriate statistical graph to present data

1. The following two statistical graphs are based on the same set of data.



Bar chart



Pie chart

Show Ans.

(a) From the bar chart / pie chart, the frequencies of different activities can be obtained more easily.

(b) From the bar chart / pie chart, the weightings of different activities can be obtained more easily.

striker 前鋒球員

2. The following frequency distribution table shows the examination scores of 40 S1D students in Chinese Language and English Language.

Chinese Language	60	70	65	55	80	50	35	22	87	96
English Language	50	55	65	70	60	70	40	40	40	88
Chinese Language	55	63	72	41	43	47	59	23	39	89
English Language	49	63	71	38	51	50	50	15	5	100
Chinese Language	71	72	70	60	5	81	62	51	48	47
English Language	69	67	75	62	30	84	65	53	48	40
Chinese Language	14	54	100	47	46	77	55	80	88	93
English Language	8	59	98	53	43	66	54	81	78	95

Consider the following statistical graphs: Show Ans.

bar chart, pie chart, broken-line graph, stem-and-leaf diagram, scatter diagram, histogram

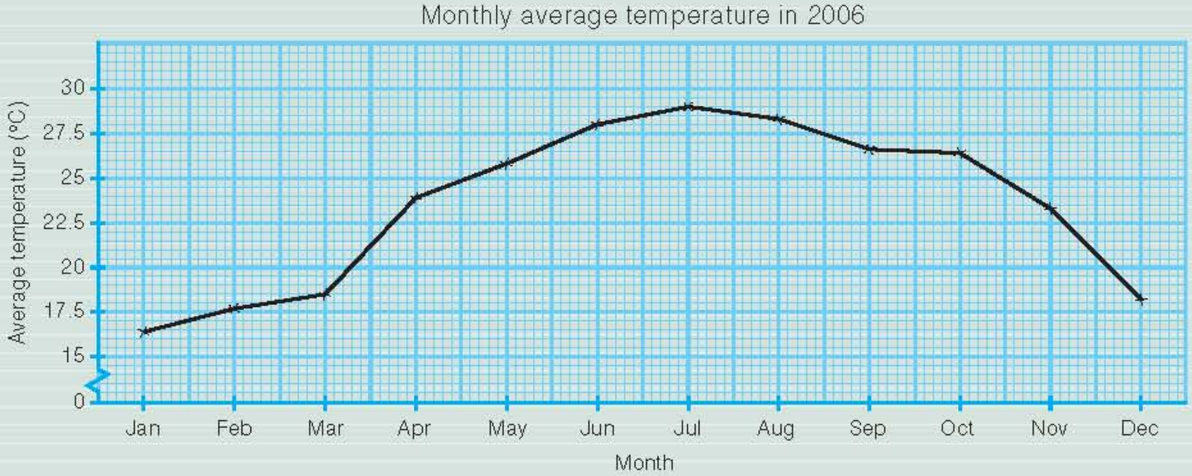
(a) Which of the above statistical graphs is more appropriate for comparing the distribution of the examination scores in the two subjects? Explain briefly.

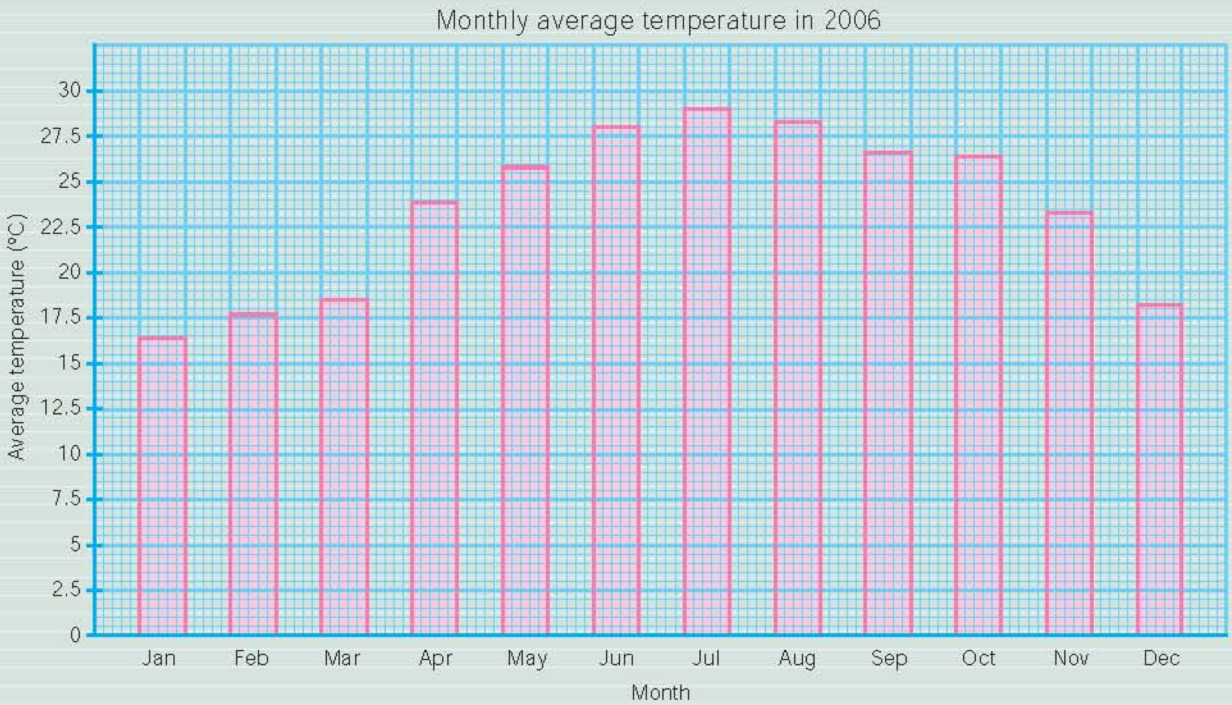
___

(b) Which of the above statistical graphs is more appropriate for studying the relation between the examination scores in the two subjects? Explain briefly.

___

3. The following graphs show the monthly average temperature recorded by the Hong Kong Observatory in 2006.





(a) From the broken-line graph / bar chart, the average temperature in October and November can be estimated more easily.

(b) Broken-line graph / Bar chart can show the continuous change in the average temperature in each month.

Show Ans.

Now I see ...

The same set of data can be presented by different statistical graphs to show different properties of the data.

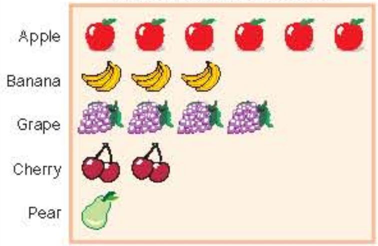
I.T. Tool

Below are some features of different statistical graphs to be referred to when choosing the appropriate graph.

1. Pictogram

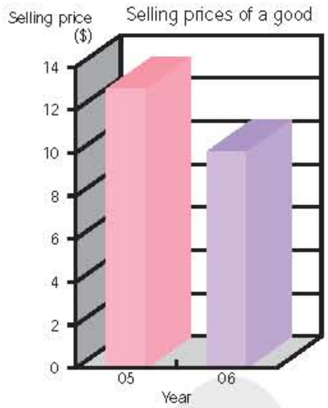
Because of the interesting pictures, it is attractive to readers and easy to understand. However, drawing is time-consuming and cannot represent numerical values accurately.

Favourite fruits with students



2. Bar chart

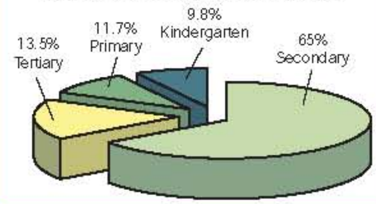
It can represent numerical values of several sets of discrete data accurately. Also, it is easy to draw.



3. Pie chart

It can represent the weight of each set of discrete data to the total data.

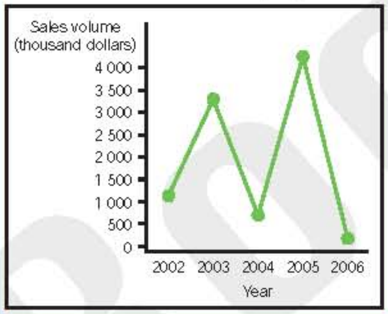
Education levels of a group of people



4. Broken-line graph

It can show the change in the data during a certain period.

Sales volume of a company



5. Stem-and-leaf diagram

It can record each datum accurately, and can be used as a tool of data management in the beginning.

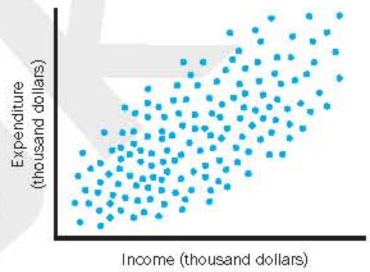
Amount spent by 30 people on lunch in a day

Stem ($10)	Leaf ($1)
1	5 6 6 8 9 9
2	0 1 2 2 4 5 5 6 6
3	0 0 2 6 6 8
4	0 5 8
5	0 8 9

6. Scatter diagram

It can show the relation between two sets of data.

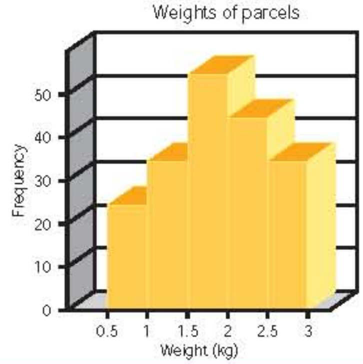
Income and expenditure



7. Histogram

It can reflect the distribution of a set of continuous data.

Weights of parcels



Extension 10.3

Solution

In each of the following situations, identify an appropriate statistical graph to present the data.

(a) Mr. Wong wants to compare his expenditure of each item with his total expenditure last month.

(b) An investor wants to observe the trend of the Hang Seng Index for the last three months.

(c) Mr. Lam wants to know the relation between the time spent by students in surfing the internet and their average examination scores.

(d) Miss Cheung wants to record all the test scores of a class in a simple and clear way.

(e) A mobile phone service provider wants to know the distribution of airtime among its customers.

(f) To show the sales records of different brands of mobile phones in a magazine.

(g) An environmental group wants to compare the water qualities among five beaches.

Chapter Summary

Chapter Summary

A. Term Introduced

[This is a quiz to check your understanding of some special terms in this chapter. Match items in column A to column B appropriately.]

Column A	Column B
1. Continuous data •	• (a) The values at the two ends of each class interval.
2. Discrete data •	• (b) Collected data which can be any values within a range.
3. Class limit •	• (c) A statistical graph which shows the relation between two sets of data.
4. Class boundary •	• (d) Collected data which are specific values.
5. Histogram •	• (e) The largest possible range of each class interval.
6. Scatter diagram •	• (f) A statistical graph which represents the frequency of each set of data by rectangular bars with no gaps in between.

B. Fact to Remember

1. $\text{Class mark} = \dfrac{\text{Lower class limit} + \text{Upper class limit}}{2}$

2. Class width = Upper class boundary – Lower class boundary

Check Yourself

[This is a quiz to remind you of the basic concepts you have learned in this chapter. Each question tests a concept under the section listed on the right. Failure in any part of a question indicates a need to do a revision on the section listed.]

Section

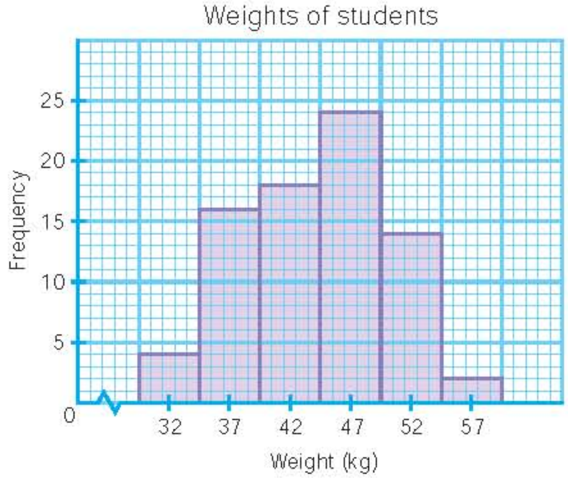
1. The following table shows the prices of different software sold in Technology Company. 10.2

Price ($)	60 - 159	160 - 259	260 - 359	360 - 459	460 - 559
Frequency	68	37	35	52	8

(a) The class interval ________________ has the highest frequency.

(b) Find the upper and lower class boundaries of the fourth class interval.

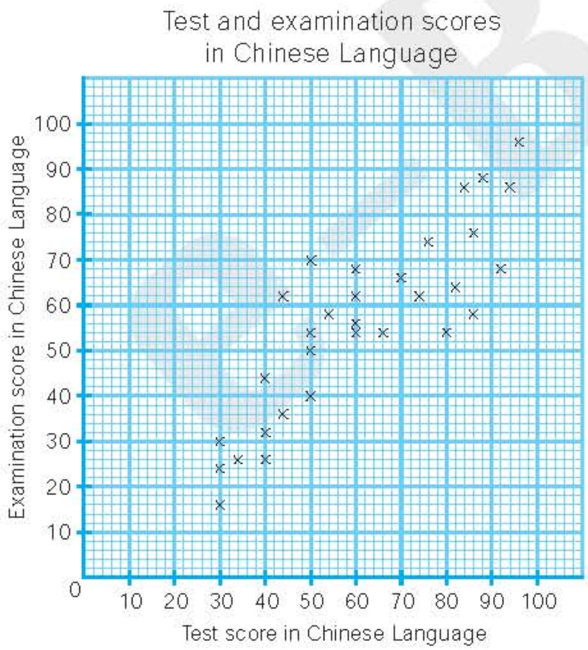
2. The following histogram shows the distribution of the weights of a group of students. It is known that the first class interval is 30 kg - 34 kg. 10.3



(a) The class interval ________________ has the fewest students.

(b) Find the number of students who weigh less than 44.5 kg.

3. The following scatter diagram shows the test and examination scores of S1B students in Chinese Language. 10.4



(a) For those students with high examination scores in Chinese Language, their test scores are high / low.

(b) How many students scored more than 80 in both the test and examination?

Revision Exercise 10

Level 1

1. The following table records the weights of 50 secondary school textbooks.

Weight (kg)	0.40 - 0.45	0.46 - 0.51	0.52 - 0.57	0.58 - 0.63	0.64 - 0.69	0.70 - 0.75
Frequency	3	11	x	7	4	2

(a) What is the value of x?

(b) Find the class mark of the second class interval.

(c) Which class interval has the lowest frequency?

2. The following table records the time taken for 30 people to run 400 m.

Time (s)	54.0 - 57.4	57.5 - 60.9	61.0 - 64.4	64.5 - 67.9	68.0 - 71.4
Frequency	4	8	7	5	6

(a) What are the upper and lower class boundaries of the third class interval?

(b) Which class interval does the class mark 66.2 s fall into?

(c) Find the number of people who took less than 60.95 s to run 400 m.

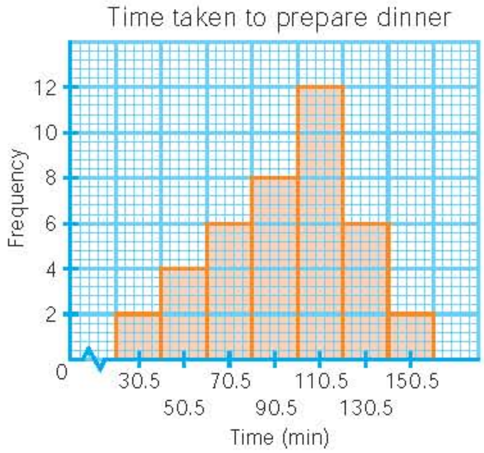
3. The following histogram shows the distribution of the heights of flowers in Beauty Garden. It is known that the first class interval is 8 cm - 12 cm.

(a) Which class interval has the highest frequency?

(b) Which class interval has the lowest frequency?

(c) How many flowers are there in Beauty Garden?

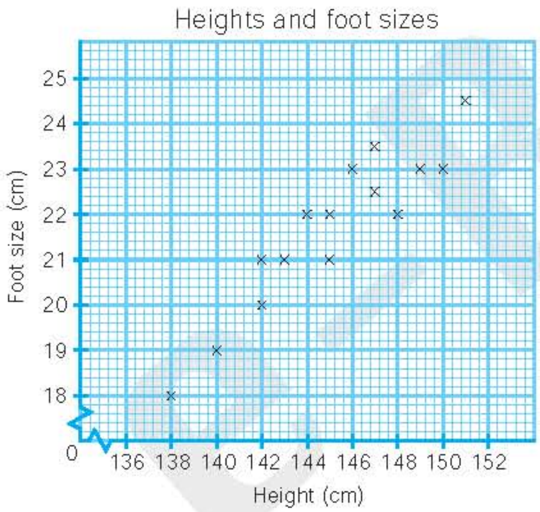
4. The following histogram shows the time taken by 40 people to prepare dinner. It is known that the first class interval is 21 min - 40 min.



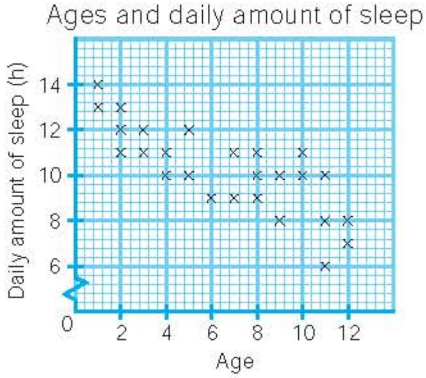
(a) Which class interval has the most people? How many people are there?

(b) What percentage of people take 120.5 min or more to prepare dinner?

5. According to the following scatter diagram, describe the relation between the heights and foot sizes of 15 students.



6. According to the scatter diagram, describe the relation between the ages of a group of children and their daily amount of sleep.



Level 2

7. The following are the collar sizes of 40 boys (in cm).

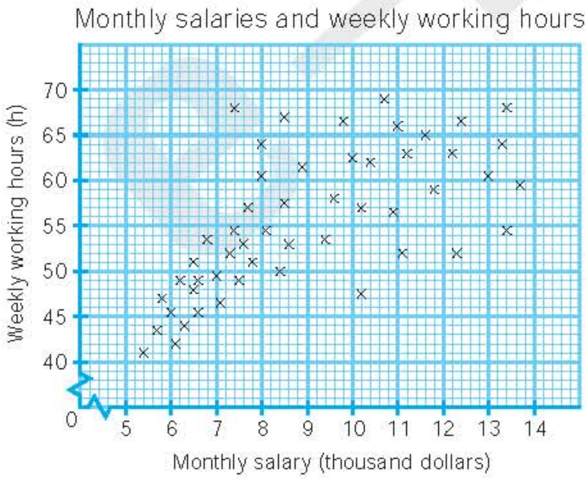
34	32	37	35	36	37	33	34
37	40	38	42	39	38	39	40
33	36	42	39	40	32	36	37
35	40	38	33	39	41	42	38
41	39	37	35	37	34	38	36

(a) According to the data above, complete the following table.

Collar size (cm)	Class boundaries (cm)	Class mark (cm)	Tally	Frequency
31 - 32				
33 - 34				
35 - 36				
37 - 38				
39 - 40				
41 - 42				

(b) Based on the table in **(a)**, draw a histogram.

8. The following scatter diagram shows the monthly salaries of a group of employees and their weekly working hours.



(a) For the employees with higher monthly salaries, are their weekly working hours also longer?

(b) For the employees with monthly salaries lower than \$8 000, describe the relation between their monthly salaries and weekly working hours.

(c) For the employees with monthly salaries higher than \$10 000, describe the relation between their monthly salaries and weekly working hours.

9. The following stem-and-leaf diagram shows the speeds of vehicles detected by the police on a highway during one day.

Speeds of vehicles

Stem (10 km/h)	Leaf (1 km/h)
7	2 3 5
8	0 0 1 2 5 8 9 9
9	0 0 1 1 2 2 2 4 4 5 6 7 7 8 9 9
10	0 0 0 1 2 5 9
11	5 6 6 9
12	0 3

(a) Using 70 km/h - 74 km/h as the first class interval, draw a frequency distribution table with the class boundaries and class mark for each class.

(b) Which class interval has the highest frequency?

(c) Suppose all the vehicles with speeds over 104.5 km/h will be fined, how many vehicles will be fined on that day? What percentage of vehicles are fined?

10. The following table shows the weights of 200 oranges.

Weight (g)	170 - 179	180 - 189	190 - 199	200 - 209	210 - 219
Frequency	20	65	35	44	36

(a) Draw a histogram to present the above data.

(b) An orange meets the standard if its weight is 199.5 g or above. Find the percentage of oranges which fail to meet the standard.

11. The following table shows the heights and weights of 10 students.

Student	Adam	Betty	Calvin	Dick	Eason	Freddy	Gary	Helen	Ivy	Joan
Height	146 cm	168 cm	158 cm	140 cm	144 cm	164 cm	150 cm	149 cm	136 cm	154 cm
Weight	56 kg	55 kg	52 kg	36 kg	39 kg	56 kg	60 kg	42 kg	33 kg	47 kg

(a) Draw a scatter diagram to present the data above.

(b) Are the taller students also heavier?

(c) How many students should watch out for their weights? Why?

12. The following are the sizes of 80 pairs of shoes sold in a shoe shop last month.

4	3	6	8	5	4	5	7	5	3	5	6	5	6	6	4	6	3	3	5
4	5	7	5	5	5	3	6	6	4	4	4	4	5	7	4	5	6	7	4
8	6	7	3	3	5	5	4	6	4	4	7	6	5	4	5	4	3	4	3
5	4	5	8	6	5	5	4	5	8	6	6	5	6	3	3	7	5	6	6

To decide how many pairs of shoes of each size are needed to be purchased next month, the owner of the shoe shop wants to draw a statistical graph using the data above for reference.

(a) What kind of statistical graph should the owner choose?

(b) Draw the statistical graph chosen in **(a)**.

MC Question

13. Which of the following must be continuous data?

I. Number of sent messages each month
II. Capacities of glasses
III. Volumes of gift boxes

A. I only
B. II only
C. I and II only
D. II and III only

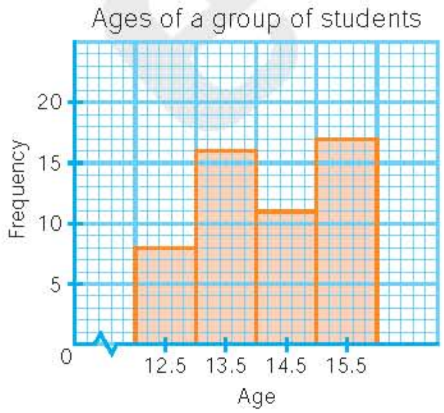
14. If the class interval is 180 - 189, what is its class width?

A. 8
B. 9
C. 10
D. 11

15. It is given that the lower class boundary of the third class interval is 18.5 kg. What is the upper class boundary of the second class interval?

A. 18.5 kg
B. 19.0 kg
C. 19.5 kg
D. It cannot be found.

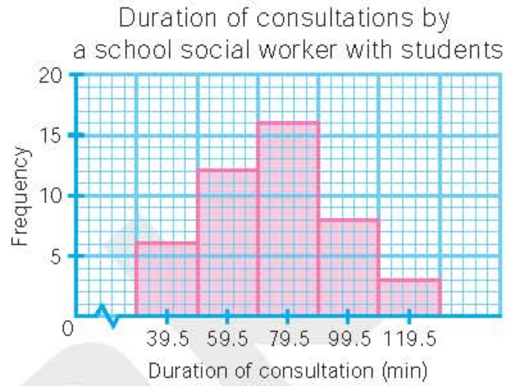
16. The histogram shows the ages of a group of students.



How many students are there in this group?

A. 17
B. 20
C. 50
D. 52

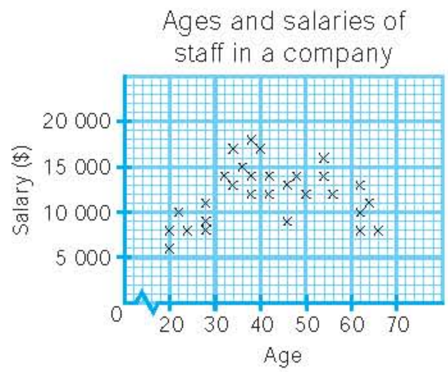
17. The following histogram shows the duration of consultations by a school social worker with students last week. It is known that the first class interval is 30 min - 49 min.



What percentage of students have consulted for 69.5 min or more?

A. 16%
B. $53\frac{1}{3}\%$
C. 60%
D. 90%

18. The scatter diagram shows the ages and salaries of staff in a company.



Staff with higher salaries are mostly between the ages of

A. 20 - 29.
B. 30 - 39.
C. 40 - 49.
D. 60 - 69.

Problem-solving and Exploring

Hint for the Title Page Question

(a) Draw a scatter diagram to present the following data.

Neck size (cm)	43.2	40.6	33.0	39.4	30.6	31.8	35.6	39.4	36.8	34.4
Waist size (cm)	86.5	81	63.5	83	61	63.5	76	77	76	70
Neck size (cm)	35.2	38.2	41.6	36	30	42	37.4	33.6	35	38.8
Waist size (cm)	68	76	80.5	72.5	60.5	85.5	73	68.5	73	74

(b) From the scatter diagram in **(a)**, describe the relation between the neck size and the waist size.

(c) How many times is the waist size equal to the neck size approximately?

(d) It is given that the neck size of Kenny is 45 cm. What is his waist size approximately?

Additional Question

1. The following table shows the average scores of a number of NBA players and their average playing time for each match.

Average score	19	23.5	22.5	8.5	2	29.5	17.5	14.5	9	3.5	33.5	27	11.5	6.5
Average playing time (min)	34	33.5	38.5	28	13	43	42.5	32	23	11.5	40.5	39	30.5	18.5
Average score	31	22	16.5	5	4.5	25	21	11.5	8	21.5	18.5	26.5	29	13.5
Average playing time (min)	37	44	39.5	15.5	14	41.5	35.5	27.5	17.5	42	37.5	43	34.5	30.5

(a) Draw a scatter diagram to present the data above.

(b) For players with average scores less than 15, describe the relation between their average scores and their average playing time for each match.

(c) For players with average playing time more than 33 min, describe the relation between their average scores and their average playing time for each match.

2. (a) Conduct a survey following the steps below.

(i) All students in your class put all the books and stationery into their school bags.

(ii) Measure the weights of each student and his/her school bag. Record the data and sex of each student.

(iii) Draw a scatter diagram each for the whole class, boys only and girls only. (Take the body weights as the x-axis, and the weights of school bags as the y-axis.)

(b) What do you discover?

I.T. Corner

A Construct a histogram

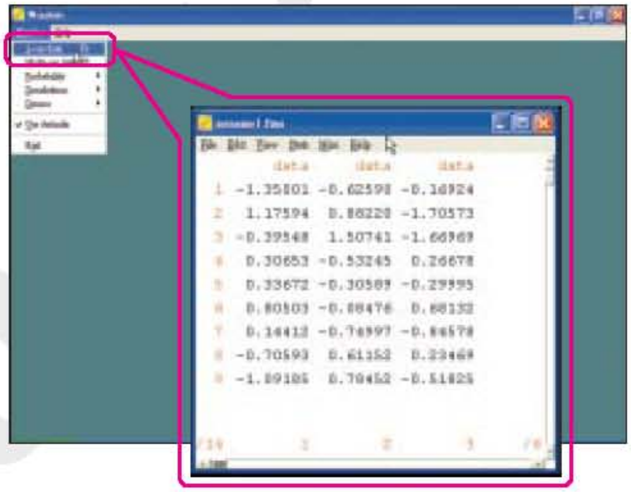
Aim: To construct a histogram according to the data in Figure 10.1 (page 10.5) of Section 10.2A

Required software: Winstats

(*Note:* Download from http://math.exeter.edu/rparris/winstats.html)

Procedures:

Step 1: Select [Window] > [1-var data].

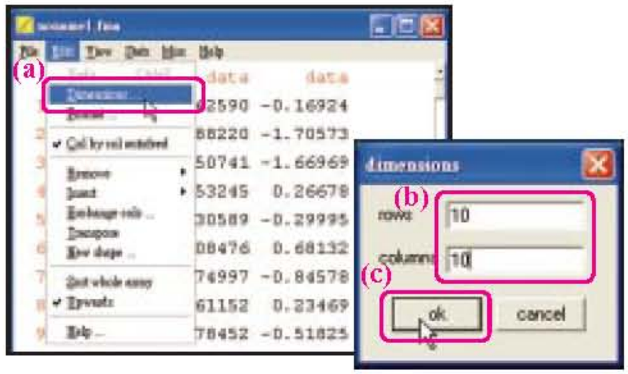


Step 2: Setting the number of data

(a) Select [Edit] > [Dimensions].

(b) Fill in the box for [rows] with '10' and that for [columns] with '10'.

(c) Click the [ok] button.

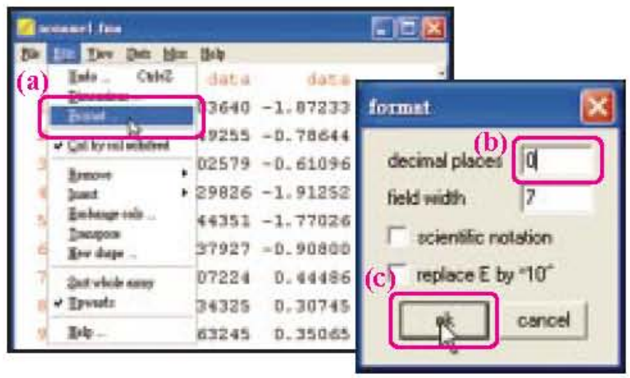


Step 3: Setting the display format of data

(a) Select [Edit] > [Format].

(b) Fill in the box for [decimal places] with '0'.

(c) Click the [ok] button.

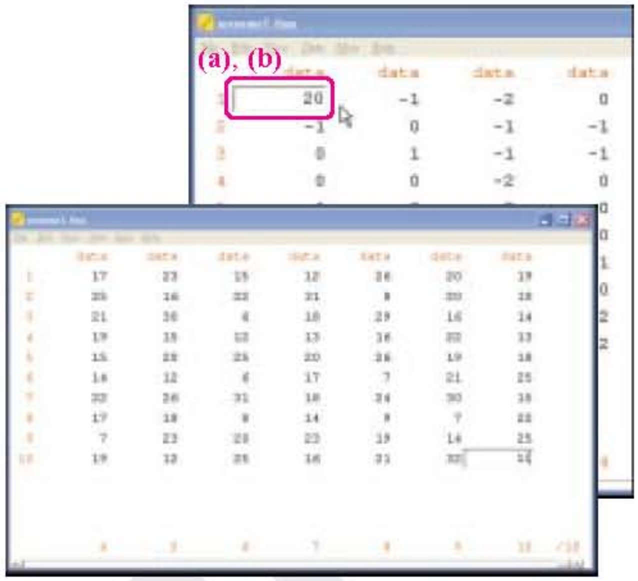


Step 4: Inputting the data

(a) Move the cursor to a location provided for inputting a datum, and then click the left button on the mouse.

(b) Type in the datum and then press Enter .

(c) Repeat **(a)** and **(b)** to input all data.

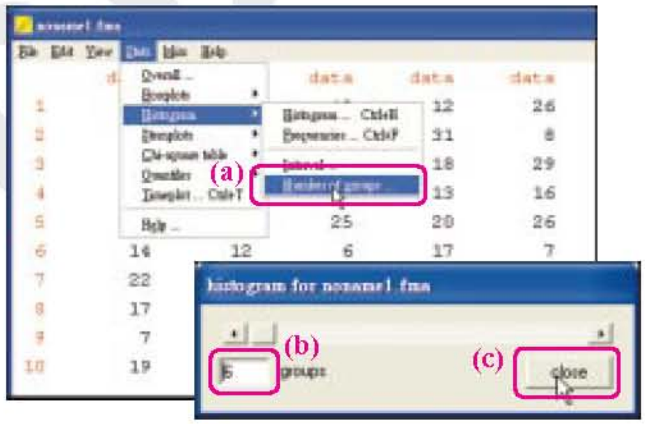


Step 5: Diagram setting and display

(a) Select [Stats] > [Histogram] > [Number of groups].

(b) Fill in the box for [groups] with '6'.

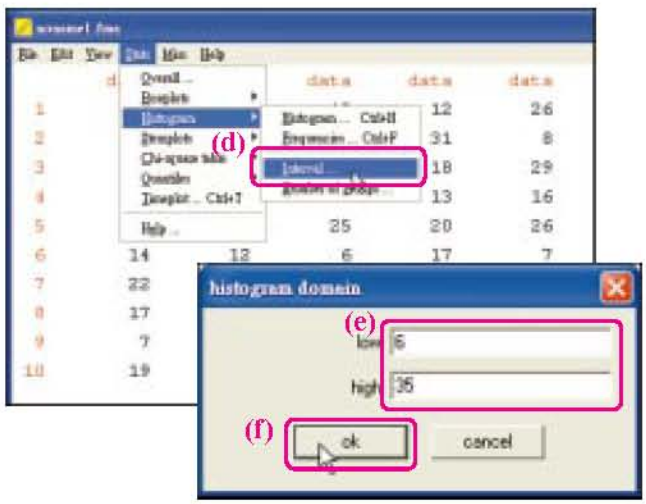
(c) Click the [close] button.



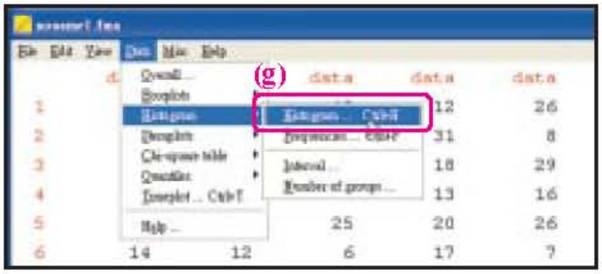
(d) Select [Stats] > [Histogram] > [Interval].

(e) Fill in the box for [low] with '6' and the box for [high] with '35'.

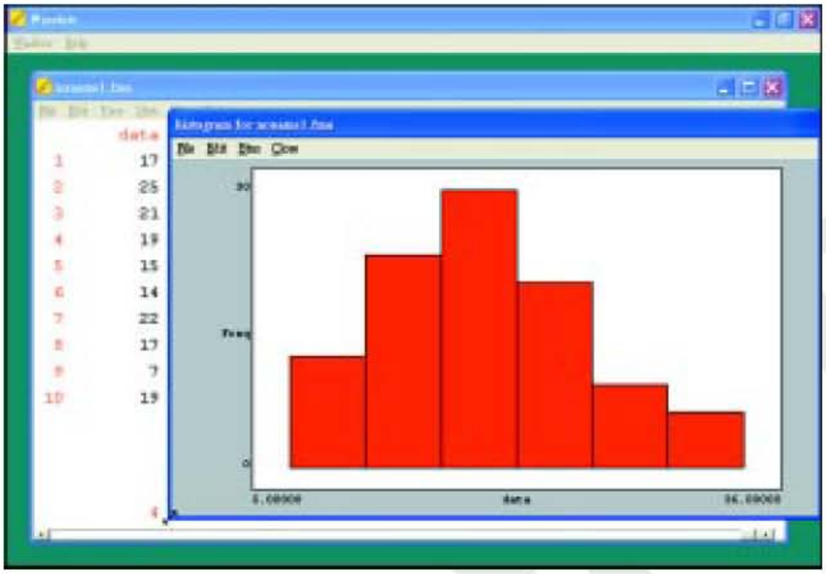
(f) Click the [ok] button.



(g) Select [Stats] > [Histogram] > [Histogram].



After completing the steps above, the result as follows will be obtained.



Practice

According to the data in question **4** (page 10.16) of Exercise 10B, draw a histogram for the average time taken for 30 people to have haircut each using computer software.

B Construct a scatter diagram

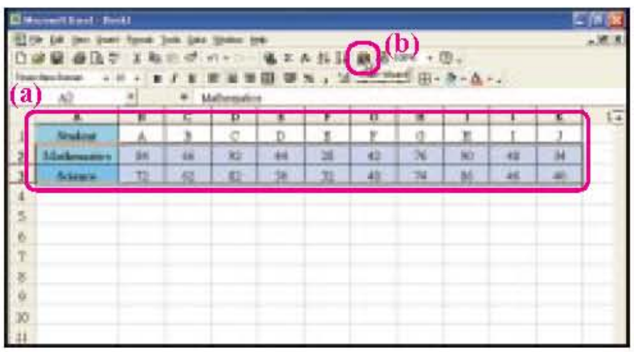
Aim: To construct a scatter diagram according to the data in Table 10.3 (page 10.17) of Section 10.4A

Required software: Microsoft Excel

Procedures:

Step 1: **(a)** Input and select the required data.

(b) Click the icon .

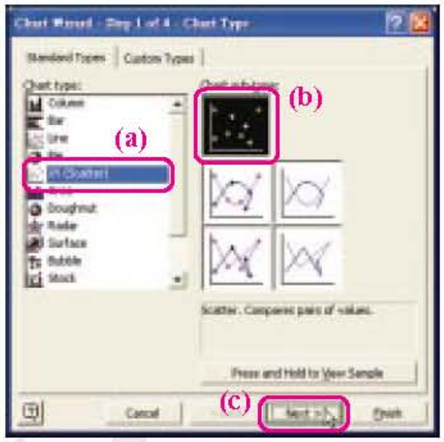


Step 2: For [Chart Type]

(a) Select [XY (Scatter)] under [Chart type].

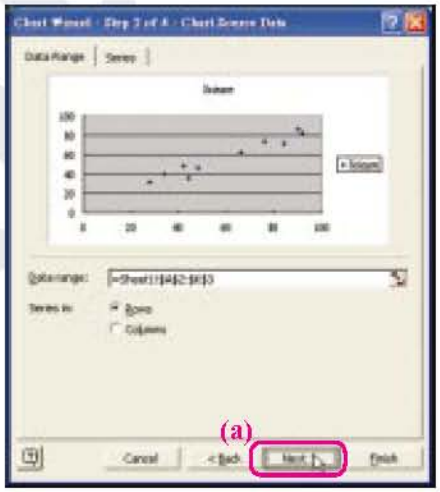
(b) Select [Scatter] under [Chart sub-type].

(c) Click the [Next] button.



Step 3: For [Chart Source Data]

(a) Click the [Next] button.



Step 4: For [Chart Options]

(a) Fill in the box for [Chart title] with 'Test scores in Mathematics and Science'; the box for [Value (X) axis] with 'Test score in Mathematics'; and the box for [Value (Y) axis] with 'Test score in Science'.

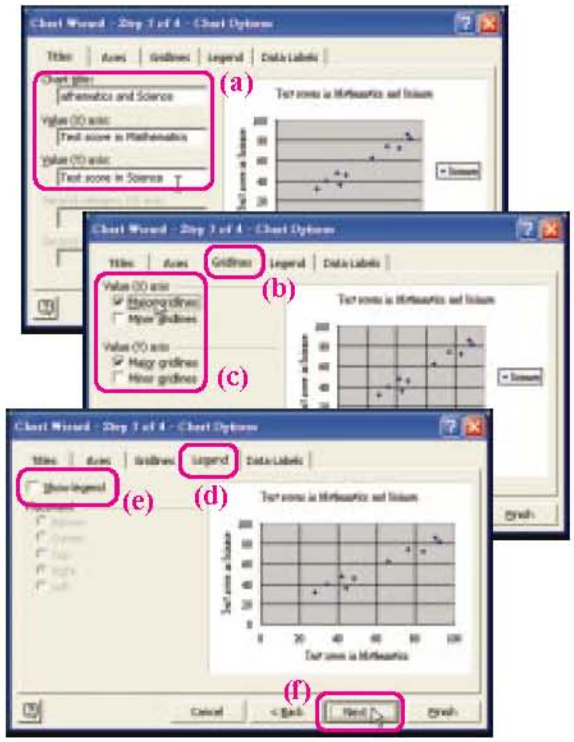
(b) Select [Gridlines].

(c) Mark the boxes [Major gridlines] for both [Value (X) axis] and [Value (Y) axis].

(d) Select [Legend].

(e) Unmark the box [Show legend] to hide the legend.

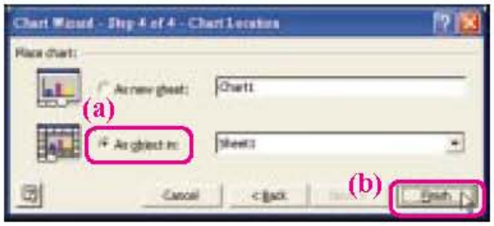
(f) Click the [Next] button.



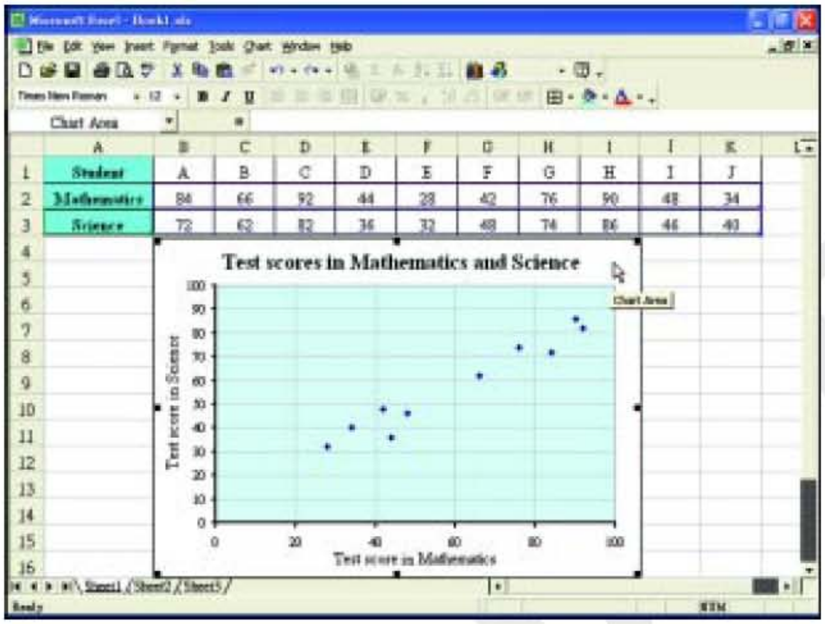
Step 5: For [Chart Location]

(a) Select [As object in].

(b) Click the [Finish] button.



After completing the steps above, the result as follows will be obtained.



Practice

According to the data in the following table, construct a scatter diagram for the test scores in Mathematics and English Language by using computer software.

Student	A	B	C	D	E	F	G	H	I	J
Mathematics	84	66	92	44	28	42	76	90	48	34
English Language	38	56	40	76	80	70	45	36	68	73